

Another Short and Elementary Proof of Strong Subadditivity of Quantum Entropy

Mary Beth Ruskai *

Department of Mathematics, Tufts University, Medford, MA 02155 USA

Marybeth.Ruskai@tufts.edu

February 1, 2008

Abstract

A short and elementary proof of the joint convexity of relative entropy is presented, using nothing beyond linear algebra. The key ingredients are an easily verified integral representation and the strategy used to prove the Cauchy-Schwarz inequality in elementary courses. Several consequences are proved in a way which allow an elementary proof of strong subadditivity in a few more lines. Some expository material on Schwarz inequalities for operators and the Holevo bound for partial measurements is also included.

1 Introduction

Because the strong subadditivity (SSA) of quantum entropy plays an important role in quantum information theory, there has been some interest in simple proofs [15, 21], suitable for elementary courses. In this note, we give a self-contained proof of SSA, valid for finite dimensional systems, using only basic linear algebra and an easily verified integral representation. The basic strategy was used in [9]. However, the presentation here, unlike that in [9] and [15], does not explicitly use the relative modular operator. Instead, the simple left and right multiplication operations explained in Section 2.1 suffice. Unlike [20, 21] not even elementary results from complex analysis are used.

*Partially supported by the National Security Agency (NSA) and Advanced Research and Development Activity (ARDA) under Army Research Office (ARO) contract number DAAD19-02-1-0065, and by the National Science Foundation under Grant DMS-0314228.

The state of a quantum system is described by a density matrix ρ , i.e., a positive semi-definite matrix satisfying $\text{Tr } \rho = 1$. The entropy of a quantum state represented by density matrix ρ was defined in 1927 by von Neumann [24, 25] as

$$S(\rho) = -\text{Tr } \rho \log \rho. \quad (1)$$

The property of SSA arises when the relevant quantum system is composed of subsystems so that ρ_{ABC} is a density matrix on a tensor product space of the form $\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$, and the partial trace is used to define the reduced density matrices, $\rho_{AB} = \text{Tr}_C \rho_{ABC}$ and $\rho_B = \text{Tr}_A \rho_{AB} = \text{Tr}_{AC} \rho_{ABC}$, etc. The SSA inequality [10] is

$$S(\rho_{ABC}) + S(\rho_B) \leq S(\rho_{AB}) + S(\rho_{BC}). \quad (2)$$

Many applications of SSA use closely related properties of the relative entropy

$$H(P, Q) = -\text{Tr } P(\log P - \log Q) \quad (3)$$

which is well-defined for positive semi-definite P, Q whenever $\ker(Q) \subset \ker(P)$ provided that we define $P(\log P - \log Q) = 0$ on $\ker(P)$. A description of the properties of $S(\rho)$ and $H(P, Q)$, and the connections between them is given in [20, 26].

The key result is the next theorem.

Theorem 1 *The relative entropy is jointly convex in P, Q , i.e., when P_j, Q_j are sequences of positive semi-definite matrices satisfying $\ker(Q_j) \subset \ker(P_j)$, then*

$$H(\sum_j x_j P_j, \sum_j x_j Q_j) \leq \sum_j x_j H(P_j, Q_j) \quad (4)$$

with $x_j \geq 0$ and $\sum_j x_j = 1$.

After proving Theorem 1 in Section 2, we obtain some important corollaries in Section 3, and show in (26) that SSA follows easily from Theorem 2b, without need for any auxiliary spaces or other results.

Although our main purpose is to present a simple proof of SSA, we added some expository material. In Section 4, we compare the argument in Section 2.4 to elementary proofs of the Cauchy-Schwarz inequality and give a direct proof of the monotonicity of relative entropy. In Section 5, we present three short proofs of the Holevo bound, each of which is valid for partial measurements.

We will frequently use expressions, such as, $A \log Q$ or $A^\dagger \frac{1}{Q} A$, without requiring the operator Q to be non-singular. But we only do so when $\ker(Q) \subset \ker(A)$ and the expression involved can be well-defined by replacing Q by $Q + \epsilon I$ and taking a limit $\epsilon \rightarrow 0+$. For simplicity and ease of exposition, we proceed as if Q is non-singular and refer to [11] for technical details.

2 Proof of joint convexity of $H(P, Q)$

2.1 Right and left multiplication

The proof will use the operations of left and right multiplication by P which are defined as $L_P(X) = PX$ and $R_P(X) = XP$. Both L_P and R_P are linear operators on the vector space of $d \times d$ matrices which becomes a Hilbert space when equipped with the Hilbert-Schmidt (HS) inner product $\langle A, B \rangle = \text{Tr } A^\dagger B$. The following properties are easy to verify

- a) The operators L_P and R_Q commute since

$$L_P[R_Q(A)] = PAQ = R_Q[L_P(A)] \quad (5)$$

even when P and Q do not commute.

- b) L_P and R_P are invertible if and only if P is non-singular, in which case $L_P^{-1} = L_{P^{-1}}$ and $R_P^{-1} = R_{P^{-1}}$.
- c) Let $\widehat{L_P}$ denote the adjoint with respect to the HS inner product. It follows from

$$\text{Tr } A^\dagger L_P(B) = \text{Tr } A^\dagger PB = \text{Tr } (P^\dagger A)^\dagger B = \text{Tr } [L_{P^\dagger}(A)]^\dagger B. \quad (6)$$

that $\widehat{L_P} = L_{P^\dagger}$ and, similarly, $\widehat{R_P} = R_{P^\dagger}$. Thus, $P = P^\dagger$ implies that the operators L_P and R_P are self-adjoint

- d) When $P \geq 0$, the operators L_P and R_P are positive semi-definite, i.e.,

$$\begin{aligned} \text{Tr } A^\dagger L_P(A) &= \text{Tr } A^\dagger P(A) \geq 0 \quad \text{and} \\ \text{Tr } A^\dagger R_P(A) &= \text{Tr } A^\dagger AP = \text{Tr } APA^\dagger \geq 0. \end{aligned}$$

2.2 Strategy

We reduce the proof of the joint convexity of $H(P, Q)$ to the proof of the following two statements.

- I) One can write the relative entropy in the form

$$H(P, Q) = \int_0^\infty \text{Tr } (Q - P) \frac{1}{L_Q + tR_P} (Q - P) \frac{1}{(1+t)^2} dt \quad (7)$$

- II) The map $(A, P, Q) \mapsto \text{Tr } A^\dagger \frac{1}{L_Q + tR_P} A$ is jointly convex in A, P, Q .

Letting $A = P - Q$ and using (II) in (I), yields the joint convexity of $H(P, Q)$.

Note that $\langle \phi, Q_j \phi \rangle > 0$ for each j implies $\langle \phi, \sum_j Q_j \phi \rangle > 0$ so that $\ker(Q_j) \subset \ker(P_j)$, implies $\ker(\sum_j Q_j) \subset \ker(\sum_j P_j)$. Thus, under the hypothesis of Theorem 1, all expressions which arise are well-defined.

2.3 Proof of the integral representation I.

We begin with the easily verified integral representation

$$-\log w = \int_0^\infty \left[\frac{1}{w+t} - \frac{1}{1+t} \right] dt \quad (8)$$

which can be rewritten as

$$= (1-w) + \int_0^\infty \frac{(w-1)^2}{w+t} \frac{1}{(1+t)^2} dt \quad (9)$$

Next, use a basis in which Q is diagonal, to see that

$$\text{Tr}(\log L_Q)(P) = \text{Tr} L_{\log Q}(P) = \text{Tr} P \log Q. \quad (10)$$

Using this and the fact that L_Q and R_P commute, one finds

$$H(P, Q) = -\text{Tr}(\log R_P^{-1})(P) - \text{Tr}(\log L_Q)(P) \quad (11)$$

$$\begin{aligned} &= -\text{Tr}[\log(L_Q R_P^{-1})](P) \\ &= \text{Tr}(1 - L_Q R_P^{-1})(P) + \\ &\quad + \int_0^\infty \text{Tr}(L_Q R_P^{-1} - 1) \frac{1}{L_Q R_P^{-1} + tI} (L_Q R_P^{-1} - 1)(P) \frac{1}{(1+t)^2} dt \end{aligned} \quad (12)$$

where the last step replaced w by $L_Q R_P^{-1}$ in (9). To see why leads to (7), first note that

$$(L_Q R_P^{-1} - 1)(P) = L_Q(I) - P = Q - P. \quad (13)$$

This can be used on the far right in (12) and also gives $\text{Tr}(1 - L_Q R_P^{-1})(P) = \text{Tr} P - Q = 0$. Next, use property (b) above to see that

$$\begin{aligned} \text{Tr} A(L_Q R_P^{-1} - 1)(B) &= \text{Tr} A(L_Q - R_P) \circ R_P^{-1}(B) \\ &= \text{Tr} [(L_Q - R_P)(A)] R_P^{-1}(B) \end{aligned} \quad (14)$$

Using this with $A = I$ and $B = (L_Q R_P^{-1} + tI)^{-1}(X)$ gives

$$\begin{aligned} \text{Tr}(L_Q R_P^{-1} - 1)(X) &= (Q - P) R_P^{-1} (L_Q R_P^{-1} + tI)^{-1}(X) \\ &= (Q - P) \frac{1}{L_Q + tR_P}(X) \end{aligned} \quad (15)$$

where we used $R_P^{-1}(L_Q R_P^{-1} + tI)^{-1} = [(L_Q R_P^{-1} + tI)R_P]^{-1} = (L_Q + tR_P)^{-1}$. Letting $X = Q - P$ and inserting (15) in (12) yields (7). **QED**

2.4 Proof of the joint convexity II:

First observe that the properties of L_P and R_Q given in Section 2.1 and the Hilbert-Schmidt inner product (6), facilitate the evaluation of such expressions as

$$\begin{aligned}\text{Tr} [(L_P + R_Q)^{-1/2}(A)]^\dagger (L_P + R_Q)^{-1/2}(B) &= \langle (L_P + R_Q)^{-1/2}(A), (L_P + R_Q)^{-1/2}(B) \rangle \\ &= \langle A, (L_P + R_Q)^{-1}(B) \rangle = \text{Tr} A^\dagger (L_P + R_Q)^{-1}(B).\end{aligned}$$

Now let $M_j = (L_{P_j} + tR_{Q_j})^{-1/2}(A_j) - (L_{P_j} + tR_{Q_j})^{1/2}(\Lambda)$, Then

$$\begin{aligned}0 &\leq \sum_j \text{Tr} M_j^\dagger M_j = \sum_j \langle M_j, M_j \rangle \\ &= \sum_j \text{Tr} A_j^\dagger (L_{P_j} + tR_{Q_j})^{-1}(A_j) - \text{Tr} (\sum_j A_j^\dagger) \Lambda \\ &\quad - \text{Tr} \Lambda^\dagger (\sum_j A_j) + \text{Tr} \Lambda^\dagger \sum_j (L_{P_j} + tR_{Q_j}) \Lambda.\end{aligned}\tag{16}$$

Next, observe that for any matrix W ,

$$\begin{aligned}\sum_j (L_{P_j} + tR_{Q_j})(W) &= \sum_j (P_j W + tW Q_j) = (\sum_j P_j) W + tW (\sum_j Q_j) \\ &= L_{\sum_j P_j}(W) + tR_{\sum_j Q_j}(W).\end{aligned}$$

Therefore, inserting the choice $\Lambda = (L_{\sum_j P_j} + tR_{\sum_j Q_j})^{-1}(\sum_j A_j)$ in (16) yields

$$\text{Tr} (\sum_j A_j)^\dagger \frac{1}{L_{\sum_j P_j} + tR_{\sum_j Q_j}} (\sum_j A_j) \leq \sum_j \text{Tr} A_j^\dagger \frac{1}{L_{P_j} + tR_{Q_j}} (A_j).\tag{17}$$

for any $t \geq 0$. Since $(xA)^\dagger \frac{1}{L_{xP} + R_{xQ}} (xA) = x \left(A^\dagger \frac{1}{L_P + R_Q} (A) \right)$ this implies¹ joint convexity. **QED**

2.5 Remarks

For simplicity, we used (11) as the starting point for obtaining the integral representation (7). It is equivalent, and customary, to begin instead with a symmetric variant of (12), $H(P, Q) = -\text{Tr} P^{1/2} [\log (L_Q R_P^{-1})] (P^{1/2})$ and then observe that $(L_Q R_P^{-1} - I)(P^{1/2}) = (R_P)^{-1/2}(Q - P)$

One advantage to our approach, like that in [15], is that it is easily extended to give a proof of joint convexity when $-\log w$ is replaced by another operator convex function. This only changes the weight function in the integral; see [9, 19] for details. Replacing $\frac{1}{(1+t)^2}$ by $\delta(1-t)$ in (7) yields $(Q - P) \frac{1}{L_P + R_Q} (Q - P)$ which is the generalized relative entropy whose Hessian yields the Riemmanian metric associated with the Bures metric $D^{\text{Bures}}(P, Q) = [2(1 - \text{Tr}(\sqrt{P}Q\sqrt{P})^{1/2})]^{1/2}$.

¹If this is not obvious, see the Appendix.

3 Consequences of joint convexity

3.1 Monotonicity of relative entropy

The joint convexity of relative entropy implies the well-known fact [13, 16, 18, 23] that it decreases under completely positive, trace-preserving (CPT) maps. These maps represent quantum channels. We will prove this result by first considering two special cases, the partial trace and the projection onto the diagonal, which are of sufficient importance to deserve separate statements and have extremely elementary proofs.

Theorem 2 *Let Φ^{QC} denote the map which projects a matrix onto its diagonal, and let Φ be any CPT map. Then*

- a) $H[\Phi^{\text{QC}}(\rho), \Phi^{\text{QC}}(\gamma)] \leq H(\rho, \gamma)$
- b) $H[\rho_A, \gamma_A] \leq H(\rho_{AB}, \gamma_{AB})$
- c) $H[\Phi(\rho), \Phi(\gamma)] \leq H(\rho, \gamma)$

Proof: First, let Z denote the diagonal unitary matrix with elements $z_{jk} = \delta_{jk}\omega^k$ with $\omega = e^{i2\pi/d}$ and note that that $(1 - \omega^{(k-n)}) \sum_j \omega^{j(k-n)} = 1 - \omega^{dj(k-n)} = 0$. Then, for any matrix X

$$\sum_j Z^j X Z^{-j} = \sum_j \omega^{j(k-n)} x_{kn} = d \delta_{kn} x_{kn} \quad (18)$$

which implies that $\Phi^{\text{QC}}(\rho) \equiv \frac{1}{d} \sum_j Z^j X Z^{-j}$ projects a matrix onto its diagonal.

Now write a bipartite state $\rho_{AB} = \sum_{jk} |e_j\rangle\langle e_k| \otimes P_{jk}$ as a block matrix with blocks P_{jk} . Then

$$H(\rho_B, \gamma_B) = H(\sum_k P_{kk}, \sum_k Q_{kk}) \leq \sum_k H(P_{kk}, Q_{kk}) \quad (19)$$

$$\begin{aligned} &= H(\sum_k |e_k\rangle\langle e_k| \otimes P_{kk}, \sum_k |e_k\rangle\langle e_k| \otimes Q_{kk}) \\ &= H[(\mathcal{I}_A \otimes \Phi^{\text{QC}})(\rho_{AB}), (\mathcal{I}_A \otimes \Phi^{\text{QC}})(\gamma_{AB})] \end{aligned} \quad (20)$$

$$\begin{aligned} &= H\left[\frac{1}{d} \sum_j (I \otimes Z)^j \rho_{AB} (I \otimes Z)^{-j}, \frac{1}{d} \sum_j (I \otimes Z)^j \gamma_{AB} (I \otimes Z)^{-j}\right] \\ &\leq \frac{1}{d} \sum_j H[(I \otimes Z)^j \rho_{AB} (I \otimes Z)^{-j}, (I \otimes Z)^j \gamma_{AB} (I \otimes Z)^{-j}] \\ &= H[\rho_{AB}, \gamma_{AB}] \end{aligned} \quad (21)$$

where Theorem 1 was used twice in the subadditive form (47), and the final equality uses the fact that conjugation of both arguments by a unitary matrix does not

change $H(\rho, \gamma)$. This proves part (b). When the space \mathcal{H}_A is 1-dimensional, the inequality between (20) and (21) yields part (a).

To prove (c) fix the ancilla representation of Lemma 3 and let

$$\sigma_{AB} \equiv U_{AB} \rho \otimes |\phi_B\rangle\langle\phi_B| U_{AB}^\dagger \quad \tau_{AB} \equiv U_{AB} \gamma \otimes |\phi_B\rangle\langle\phi_B| U_{AB}^\dagger.$$

Then $\Phi(\rho) = \sigma_A$, $\Phi(\gamma) = \tau_A$ and, since U_{AB} is unitary, $H(\rho, \gamma) = H(\sigma_{AB}, \tau_{AB})$. Thus, it follows from part (b) that

$$H[\Phi(\rho), \Phi(\gamma)] = H(\sigma_A, \tau_A) \leq H(\sigma_{AB}, \tau_{AB}) = H(\rho, \gamma). \quad \text{QED} \quad (22)$$

3.2 Convexity corollaries

The conditional entropy is given by

$$S(\rho_{AB}) - S(\rho_A) = -H(\rho_{AB}, \rho_A \otimes \tfrac{1}{d}I) + \log d, \quad (23)$$

It then follows immediately from the joint convexity of $H(\rho, \gamma)$ that

$$\rho_{AB} \mapsto S(\rho_{AB}) - S(\rho_A) \quad \text{is concave.} \quad (24)$$

Moreover, for any CPT map Φ , the map

$$\rho \mapsto S(\rho) - S[\Phi(\rho)] \quad \text{is concave.} \quad (25)$$

This follows from (24). Use the same notation as in part (c) of the previous section and observe that $S(\rho) - S[\Phi(\rho)] = S(\sigma_{AB}) - S(\sigma_B)$.

3.3 Completing the proof of SSA

The SSA inequality (2) follows immediately from Corollary 2b with $\gamma = \rho_{AC} \otimes \tfrac{1}{d}I$. We write this out explicitly using (23).

$$\begin{aligned} S(\rho_A) - S(\rho_{AB}) &= H(\rho_{AB}, \rho_A \otimes \tfrac{1}{d}I) - \log d \\ &\leq H(\rho_{ABC}, \rho_{AC} \otimes \tfrac{1}{d}I) - \log d \\ &= S(\rho_{ABC}) - S(\rho_{AB}). \end{aligned} \quad \text{QED} \quad (26)$$

There is another form of SSA which follows easily from (24), namely,

$$S(\rho_B) + S(\rho_D) \leq S(\rho_{AB}) + S(\rho_{AD}). \quad (27)$$

To prove this first consider

$$F(\rho_{ABD}) = S(\rho_{AB}) + S(\rho_{AD}) - S(\rho_B) - S(\rho_D). \quad (28)$$

When ρ_{ABD} is pure, it follows from Lemma 4 that $S(\rho_{AB}) = S(\rho_D)$ and $S(\rho_{AD}) = S(\rho_B)$. Thus, $F(\rho_{ABD}) = 0$ for pure states. Since $F(\rho_{ABC})$ is the sum of two functions $S(\rho_{AB}) - S(\rho_B)$ and $+S(\rho_{AD}) - S(\rho_D)$ which are concave by (24), the map $\rho_{ABD} \mapsto F(\rho_{ABD})$ is also concave. Since any mixed ρ_{ABD} is a convex combination of pure states, $F(\rho_{ABD}) \geq 0$, which implies (27).

By Lemma 5, one can purify ρ_{ABC} or ρ_{ABD} to ρ_{ABCD} and use Lemma 4 to show that (27) holds if and only if (2) does.

4 Remarks on Cauchy-Schwarz type inequalities

4.1 Elementary proof strategy

The elementary vector version of the Cauchy-Schwarz inequality states that

$$\left| \sum_k \bar{v}_k w_k \right|^2 \leq \left(\sum_k |v_k|^2 \right) \left(\sum_k |w_k|^2 \right) \quad (29)$$

When $v_k = p_k^{1/2}$, $w_k = p_k^{-1/2} a_k$, this can be written as

$$\sum_k \bar{a}_k \frac{1}{\sum_k p_k} \sum_k a_k \leq \sum_k \bar{a}_k \frac{1}{p_k} a_k. \quad (30)$$

In [11], Lieb and Ruskai proved an operator version of (30), namely that

$$\sum_k A_k^\dagger \frac{1}{\sum_k P_k} \sum_k A_k \leq \sum_k A_k^\dagger \frac{1}{P_k} A_k. \quad (31)$$

holds as an operator inequality. This is equivalent to the statement that the map $(A, P) \mapsto A^\dagger P^{-1} A$ is jointly operator convex. The proof in Section 2.4 is based on that in [11] which (although published later) actually preceded the proof of SSA. However, without the additional ingredient of L_P and R_Q , which are motivated by Araki's subsequent introduction [1] of the relative modular operator, the results in [11] are not sufficient to prove SSA. The recognition that the argument in [11] could be modified to prove SSA took another 25 years [9].

The proofs in both [11] and Section 2.4 are variants of the standard strategy used to prove the elementary inequality (29). One observes that $\sum_k |v_k + \lambda w_k|^2 \geq 0$ and shows that the minimizing choice $\lambda = -(\sum_k v_k) / (2 \sum_k w_k)$ yields (29). In Section 2.4 the operator Λ plays the role of λ .

4.2 Schwarz inequalities for CP maps

We now consider Schwarz type inequalities involving completely positive (CP) maps. Let Φ be a CP map written in Kraus form $\Phi(P) = \sum_j K_j P K_j^\dagger$. It then follows from (31) that

$$\begin{aligned} [\Phi(A)]^\dagger \frac{1}{\Phi(P)} \Phi(A) &= \sum_j K_j A^\dagger K_j^\dagger \frac{1}{\sum_j K_j P K_j^\dagger} \sum_j K_j A K_j^\dagger \\ &\leq \sum_j K_j A^\dagger \frac{1}{P} A K_j^\dagger = \Phi\left(A^\dagger \frac{1}{P} A\right). \end{aligned} \quad (32)$$

By making the replacements $A \rightarrow B^\dagger A$ and $P \rightarrow B^\dagger B$ in (32) one finds

$$\Phi(A^\dagger B) \frac{1}{\Phi(B^\dagger B)} \Phi(B^\dagger A) \leq \Phi(A^\dagger A) \quad (33)$$

This inequality is proved in [11] using the Stinespring [17, 22] representation..

Choi [3] realized that (32) and (33) hold under the weaker condition that Φ is 2-positive. His approach is quite different, and based on the fact that a 2×2 block matrix $\begin{pmatrix} P & C \\ C^\dagger & Q \end{pmatrix}$ is positive semi-definite if and only if $C = \sqrt{P} X \sqrt{Q}$ with X a contraction, i.e., $X^\dagger X \leq I$. When P, Q are both non-singular, this is equivalent to $C^\dagger P^{-1} C \leq Q$. The 2-positivity of Φ says that $\begin{pmatrix} \Phi(A^\dagger A) & \Phi(A^\dagger B) \\ \Phi(B^\dagger A) & \Phi(B^\dagger B) \end{pmatrix}$ is positive semi-definite. Applying the condition above yields (33).

4.3 Monotonicity of relative entropy

In [9] a strategy similar to that in Section 2.4 was used to give a direct proof of the monotonicity of relative entropy under CPT maps without using an auxiliary space. One begins as before, but with $M = (L_P + tR_Q)^{-1/2}(A) - (L_P + tR_Q)^{1/2}[\hat{\Phi}(X)]$ and $X = (L_{\Phi(P)} + tR_{\Phi(Q)})^{-1}[\Phi(A)]$. Then $M^\dagger M \geq 0$ implies

$$\begin{aligned} \text{Tr } A^\dagger \frac{1}{L_P + tR_Q} A - 2\text{Tr } [\Phi(A)]^\dagger \frac{1}{L_{\Phi(P)} + tR_{\Phi(Q)}} \Phi(A) \\ + \text{Tr } [\hat{\Phi}(X)]^\dagger [L_P + tR_Q] \hat{\Phi}(X) \geq 0 \end{aligned} \quad (34)$$

Comparing the last two terms requires a bit more work and the use of (33). Since Φ trace preserving implies $\hat{\Phi}(I) = I$, (33) implies $[\hat{\Phi}(X)]^\dagger \hat{\Phi}(X) \leq \hat{\Phi}(X^\dagger X)$ and

$\widehat{\Phi}(X)[\widehat{\Phi}(X)]^\dagger \leq \widehat{\Phi}(XX^\dagger)$. Then, using the cyclicity of the trace, one finds

$$\begin{aligned} \text{Tr} [\widehat{\Phi}(X)]^\dagger [L_P + tR_Q] \widehat{\Phi}(X) &= \text{Tr} \widehat{\Phi}(X) [\widehat{\Phi}(X)]^\dagger P + t \text{Tr} [\widehat{\Phi}(X)]^\dagger \widehat{\Phi}(X) Q \\ &\leq \text{Tr} [\widehat{\Phi}(XX^\dagger)P + t\widehat{\Phi}(X^\dagger X)Q] \end{aligned} \quad (35)$$

$$\begin{aligned} &= \text{Tr} [XX^\dagger \Phi(P) + tX^\dagger X \Phi(Q)] \\ &= \text{Tr} X^\dagger [L_{\Phi(P)} + tR_{\Phi(Q)}] X \\ &= [\Phi(A)]^\dagger \frac{1}{L_{\Phi(P)} + tR_{\Phi(Q)}} \Phi(A) \end{aligned} \quad (36)$$

Using this in (34) allows one to combine the last two terms as before. Substituting the resulting inequality in (7) yields part (c) of Theorem 2.

5 Holevo bounds for partial measurements

In order to state the Holevo bound, we introduce some notation. Let \mathcal{E} denote an ensemble $\{\pi_j, \rho_j\}$ with $\pi_j > 0$, $\sum_j \pi_j = 1$ and each ρ_j a density matrix. The Holevo χ -quantity is defined as

$$\chi(\mathcal{E}) = S\left(\sum_j \pi_j \rho_j\right) - \sum_j \pi_j S(\rho_j). \quad (37)$$

A set of positive semi-definite operators $\{M_a\}$ satisfying $\sum_a M_a$ is called a positive operator valued measurement (POVM) and denoted \mathcal{M} . Every POVM defines a CPT map $\Phi_{\mathcal{M}}$ which takes $\rho \mapsto \sum_a (\text{Tr} \rho M_a) |a\rangle\langle a|$. The usual Holevo bound states that

$$\chi(\mathcal{E}) \geq \chi[\Phi_{\mathcal{M}}(\mathcal{E})] \equiv S\left[\sum_j \pi_j \Phi_{\mathcal{M}}(\rho_j)\right] - \sum_j \pi_j \Phi_{\mathcal{M}}(\rho_j). \quad (38)$$

where $\Phi_{\mathcal{M}}(\mathcal{E})$ denotes the ensemble in which each ρ_j is replaced by $\Phi_{\mathcal{M}}(\rho_j)$.

There is now an extensive literature on bounds involving partial measurements. Consider the situation in which two parties, Alice and Bob, share an ensemble of (possibly entangled) states $\{\pi_j, \rho_j^{AB}\}$ on $\mathcal{H}_A \otimes \mathcal{H}_B$, on which one of the parties makes a measurement. In such cases, one expects a bound of the form

$$\chi(\mathcal{E}^{AB}) \geq \chi[(I \otimes \Phi_{\mathcal{M}_B})(\mathcal{E}^{AB})] \geq \chi[(\Phi_{\mathcal{M}_A} \otimes \Phi_{\mathcal{M}_B})(\mathcal{E}^{AB})]. \quad (39)$$

We observe that three simple strategies for proving (38) easily extend to (39).

The first proof uses the observation of Yuen and Ozawa [27] that

$$S(\rho_{\text{av}}) - \sum_j \pi_j S(\rho_j) = \sum_j \pi_j H(\rho_j, \rho_{\text{av}}). \quad (40)$$

where $\rho_{\text{av}} = \sum_j \pi_j \rho_j$. It follows from part (c) of Theorem 2 that

$$\begin{aligned} & H[(\Phi_{\mathcal{M}_A} \otimes \Phi_{\mathcal{M}_B})(\rho_j^{AB}), (\Phi_{\mathcal{M}_A} \otimes \Phi_{\mathcal{M}_B})(\rho_{\text{av}}^{AB})] \\ & \leq H[I \otimes \Phi_{\mathcal{M}_B})(\rho_j^{AB}), (I \otimes \Phi_{\mathcal{M}_B})(\rho_{\text{av}}^{AB})] \leq H(\rho_j^{AB}, \rho_{\text{av}}^{AB}) \end{aligned} \quad (41)$$

which is equivalent to (39).

The next proof uses the fact that $\chi(\mathcal{E})$ can be regarded as a form of mutual information between the quantum states ρ_j and their classical probability distribution π_j . Let $\gamma_{QC} = \sum_j \pi_j \rho_j \otimes |j\rangle\langle j|$ be a density matrix on $\mathcal{H}_Q \otimes \mathcal{H}_C$. Then, as was observed in [6],

$$\chi(\mathcal{E}) = S(\rho_{\text{av}}) - \sum_j \pi_j S(\rho_j) = H(\gamma_{QC}, \gamma_Q \otimes \gamma_C). \quad (42)$$

Then part (c) of Theorem 2 gives

$$H[(\Phi \otimes I)(\gamma_{QC}), (\Phi \otimes I)(\gamma_Q \otimes \gamma_C)] \leq H(\gamma_{QC}, \gamma_Q \otimes \gamma_C). \quad (43)$$

which is equivalent to (38). To obtain (39), let $\mathcal{H}_Q = \mathcal{H}_A \otimes \mathcal{H}_B$ and observe that

$$\begin{aligned} & H[(\Phi \otimes \Phi \otimes I)(\gamma_{ABC}), (\Phi \otimes \Phi \otimes I)(\gamma_{AB} \otimes \gamma_C)] \\ & \leq H[(\Phi \otimes I \otimes I)(\gamma_{ABC}), (\Phi \otimes I \otimes I)(\gamma_{AB} \otimes \gamma_C)] \\ & \leq H(\gamma_{ABC}, \gamma_{AB} \otimes \gamma_C). \end{aligned} \quad (44)$$

The final proof uses the observation in [12], that the Holevo bound (37) is equivalent to the statement that $\rho \mapsto S(\rho) - S[\Phi_{\mathcal{M}}(\rho)]$ is convex, which is a special case of (25). Thus, the bound (39) follows immediately from (25) with Φ replaced first by $I_A \otimes \Phi_{\mathcal{M}_B}$ and then by $\Phi_{\mathcal{M}_A} \otimes \Phi_{\mathcal{M}_B}$.

A Appendix

Let $A = \sum_k \lambda_k |\phi_k\rangle\langle\phi_k|$ be a self-adjoint matrix with eigenvalues λ_k in the domain of the function $f(w)$. Then we define $f(A) = \sum_k f(\lambda_k) |\phi_k\rangle\langle\phi_k|$. This is equivalent to any other reasonable definition and implies that substituting $L_Q R_P^{-1}$ for w to obtain (12) is fully justified; there is no need to explicitly find the eigenvalues and eigenvectors of $L_Q R_P^{-1}$.

If a function F satisfies $F(xA) = xF(A)$ then convexity is equivalent to subadditivity. First, observe that when F is also convex

$$\frac{1}{2}F(A+B) = g\left(\frac{1}{2}[A+B]\right) \leq \frac{1}{2}F(A) + \frac{1}{2}F(B). \quad (45)$$

Conversely, if F is subadditive, then

$$F[xA + (1-x)B] \leq F(xA) + F[(1-x)B] = xF(A) + (1-x)F(B). \quad (46)$$

Although the relative entropy $H(P, Q)$ is usually considered for density matrices, (3) defines it more broadly. Since Klein's inequality [14, 16, 20] says that $H(P, Q) \geq \text{Tr } P - \text{Tr } Q$, it follows that $H(P, Q) \geq 0$ when $\text{Tr } P = \text{Tr } Q$. It is easy to verify that $H(xP, xQ) = xH(P, Q)$ for $x > 0$. Therefore, by the observations above, (4) is equivalent to

$$H(\sum_j P_j, \sum_j Q_j) \leq \sum_j H(jP_j, Q_j). \quad (47)$$

For completeness, we also state some well-known results used in proving corollaries to the joint convexity and SSA. None are needed to obtain a proof of SSA.

Lemma 3 (Ancilla representation) *Any CPT map $\Phi : M_d \mapsto M_d$ can be represented using an auxiliary space \mathcal{H}_B as*

$$\Phi(\rho) = \text{Tr}_B U_{AB} \rho \otimes |\phi_B\rangle\langle\phi_B| U_{AB}^\dagger \quad (48)$$

where U_{AB} is unitary and $|\phi_B\rangle\langle\phi_B|$ is a pure state. If $\sigma_{AB} \equiv U_{AB} \rho \otimes |\phi_B\rangle\langle\phi_B| U_{AB}^\dagger$, then $\Phi(\rho) = \gamma_A$ and $S(\sigma_{AB}) = S(\rho)$.

This is essentially a corollary to the Stinespring representation theorem [22]. It was introduced in the form mused here by Lindblad [13] who made the observation about entropy and used it to give the first proof of Theorem 2c. For an overview of representation theorems, see Chapter 2 of Paulsen [17]; for short accessible summaries, see the appendices to [4, 7, 8] as well as Section III.D of [20]. The term ‘‘ancilla representation’’ is introduced in [8].

The following well-known, and easily proved, facts go back at least to [2]. For further references and discussion see [5, 14, 20].

Lemma 4 *When $\rho_{AB} = |\psi_{AB}\rangle\langle\psi_{AB}|$ is a pure state, its reduced density matrices ρ_A and ρ_B have the same non-zero eigenvalues and $S(\rho_A) = S(\rho_B)$.*

Lemma 5 *Given a density matrix ρ in M_d of rank m , one can find a pure ρ_{AB} in $M_d \otimes M_m$ with $\rho_A = \rho$.*

References

- [1] H. Araki, “Relative Entropy of State of von Neumann Algebras” *Publ RIMS Kyoto Univ.* **9**, 809–833 (1976).
- [2] H. Araki and E. Lieb, “Entropy inequalities” *Comm. Math. Phys.* **18**, 160–170 (1970).
- [3] M-D. Choi, “A Schwarz inequality for positive linear maps on C^* -algebras” *Illinois J. Math.* **18**, 565–574 (1974).
- [4] A. S. Holevo, “On complementary channels and the additivity problem” quant-ph/0509101
- [5] C. King and M.B. Ruskai, “Minimal Entropy of States Emerging from Noisy Quantum Channels” *IEEE Trans. Info. Theory* **47**, 192–209 (2001).
- [6] C. King and M.B. Ruskai, “Capacity of Quantum Channels Using Product Measurements” *J. Math. Phys.* **42**, 87–98 (2001).
- [7] C. King, K. Matsumoto, M. Nathanson and M. B. Ruskai, “Properties of Conjugate Channels with Applications to Additivity and Multiplicativity” quant-ph/0509126
- [8] D. Kretschmann and R. F. Werner “Quantum Channels with Memory” quant-ph/0502106
- [9] A. Lesniewski and M.B. Ruskai, “Monotone Riemannian metrics and relative entropy on non-commutative probability spaces” *J. Math. Phys.* **40**, 5702–5723 (1999).
- [10] E.H. Lieb and M.B. Ruskai, “Proof of the Strong Subadditivity of Quantum Mechanical Entropy” *J. Math. Phys.* **14**, 1938–1941 (1973).
- [11] E.H. Lieb and M.B. Ruskai, “Some Operator Inequalities of the Schwarz Type” *Adv. Math* **12**, 269–273 (1974).
- [12] E.H. Lieb and R. Seiringer “A Stronger Subadditivity of Entropy” *Phys. Rev. A* **71**, 062329 (2005). math-ph/0412009
- [13] G. Lindblad “Completely Positive Maps and Entropy Inequalities” *Commun. Math. Phys.* **40**, 147–151 (1975).
- [14] M. Nielsen and I. Chuang *Quantum Computation and Quantum Information* (Cambridge University Press, 2000).

- [15] M. A. Nielsen and D. Petz “A simple proof of the strong subadditivity inequality” *QIC* quant-ph/0408130
- [16] M. Ohya and D. Petz, *Quantum Entropy and Its Use* (Springer-Verlag, 1993; second edition, 2004).
- [17] V. Paulsen, *Completely Bounded Maps and Operator Algebras* (Cambridge University Press, 2002).
- [18] D. Petz, “Quasi-Entropies for Finite Quantum Systems” *Rep. Math. Phys.* **23**, 57–65 (1986).
- [19] D. Petz, “Monotone Metrics on Matrix spaces”, *Lin. Alg. Appl.* **244**, 81–96 (1996).
- [20] M.B. Ruskai, “Inequalities for Quantum Entropy: A Review with Conditions for Equality” *J. Math. Phys.* **43**, 4358–4375 (2002); erratum **46**, 019901 (2005).
- [21] M.B. Ruskai, Lieb’s simple proof of concavity of $\text{Tr } A^p K^\dagger B(1-p)K$ and remarks on related inequalities” quant-ph/0404126
- [22] W.F. Stinespring, “Positive functions on C^* -algebras” *Proc. Amer. Math. Soc.* **6**, 211–216 (1955).
- [23] A. Uhlmann, “Relative Entropy and the Wigner-Yanase-Dyson-Lieb Concavity in an Interpolation Theory” *Commun. Math. Phys.* **54**, 21–32 (1977).
- [24] J. von Neumann, “Thermodynamik Quantenmechanischer Gesamtheiten” *Gött. Nach.* **1**, 273–291 (1927).
- [25] J. von Neumann *Mathematische Grundlagen der Quantenmechanik* (Berlin, 1932); English translation by R.T. Beyer *Mathematical Foundations of Quantum Mechanics* (Princeton, 1955).
- [26] A. Wehrl “General Properties of Entropy” *Rev. Mod. Phys.* **50** 221–260 (1978).
- [27] H. P. Yuen and M. Ozawa “Ultimate information carrying limit of quantum systems” *Phys. Rev. Lett.* **70**, 363–366 (1993).